CSE 2194: Supervised Machine Learning
Programming Assignment-II
(Data Visualization: part-1)

Question 1:

Create Dataframes using pandas using the given dictionary. The data name should be person_details.csv. Save the data in your local directory and print the first five data entities (rows).

```
data = {
   'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Emily', 'Frank', 'Grace', 'Hannah', 'Isaac', 'Jack'],
   'Gender': ['Female', 'Male', 'Male', 'Male', 'Female', 'Male', 'Female', 'Female', 'Male', 'Male'],
   'Age': [25, 30, 35, 40, 45, 50, 55, 60, 65, 70],
   'City': ['New York', 'Los Angeles', 'Chicago', 'Houston', 'Phoenix', 'Philadelphia', 'San Antonio',
'San Diego', 'Dallas', 'San Jose'],
   'Living Expenses': [3000, 3500, 3200, 2800, 4000, 3300, 3100, 3700, 2900, 3800]
}
```

Question 2:

Import the data (person_details.csv) created in question 1, add two more data entities with details of your friend and yours, and print the last five rows.

Question 3:

Import the data (person_details.csv) and perform the following tasks

A. Use iloc to select the first three rows and the first two columns of the DataFrame.
B. Use loc to select all the data of individuals whose living expenses are greater than 3500.
C. Use loc to select the names and ages of individuals who are females and live in either New York or Los Angeles.
D. Use iloc to select the rows from index 2 to index 7 (inclusive) and all columns.

Question 4:

A two-dimensional NumPy array representing student scores in different subjects. Each row represents a student, and each column represents a subject. Write Python code to the following tasks:

A. Calculate the average score for each subject.
B. Determine which student has the highest average score across all subjects.
C. Find the subject in which the highest-scoring student has scored the lowest.

```
scores = [
   [80, 90, 70, 85, 95],
   [75, 85, 95, 70, 80],
   [85, 95, 75, 80, 90],
   [90, 80, 85, 75, 70]
]
```

Question 5:

Given the heights (in inches) of two groups of students, Group A and Group B. Write a Python program to determine the difference in the mean heights between the two groups.
group_a_heights = [65, 68, 70, 72, 64, 67, 71, 66, 68, 69]
group_b_heights = [62, 64, 67, 68, 63, 65, 66, 61, 64, 67]

Question 6:

A group of 8 students is forming a committee of 3 members. Write a Python program using SciPy to calculate the total number of possible committees that can be formed.

Question 7:

Write a Python program to calculate the determinant and inverse of a given matrix. (Use SciPy)
A = [[3, 2, -1],
[2, -4, 7],
[1, 1, 1]]

Question 8:

Given a company's monthly sales data (in thousands of dollars) for the past year. The data is stored in a Python list. Write Python code using Matplotlib to create a line plot to visualize the trend of monthly sales over the past year. Add labels to the x-axis and y-axis with appropriate names. Add a title to the plot indicating what it represents. Save the image in your local directory.
monthly_sales = [50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105]
(Hints: The x-axis is months [1-12], and the y-axis is monthly sales)

Question 9:

Given a dataset representing the performance scores of students in a class, Write Python code using Matplotlib to perform the following tasks:

A. Create a bar graph showing the SML scores of each student.
B. Create a histogram showing the distribution of AD scores.
C. Create a scatter plot to visualize the relationship between SML and AD scores.
D. Create a pie chart to represent the distribution of scores across students.

Students	SML_scores	AD_score
Alice	85	80
Bob	90	75
Charlie	70	85
David	80	70
Emily	75	90

Question 10:

A dataset containing information about a company's monthly sales revenue (in thousands of dollars) for the past year, along with the corresponding months. Write Python code using Seaborn to create a line plot to visualize the trend of monthly sales revenue over the past year.

```
data = {
   'Month': ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec'],
   'Sales_Revenue': [50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105]
}
```

——–——----------**********----------